Shotgun Library Utilization for Microbial Sequencing Projects at the JGI

Chris Detter, Eileen Dalin, Doug Smith, Hope Tice, Mariana Anaya, Dean Ng, Victor Dorsett, Alex Copeland, Kerrie Barry, Miranda Smith-Harmon, Susan Lucas, Eddy Rubin, and Paul Richardson.

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

To date the US Department of Energy's Joint Genome Institute has produced shotgun libraries for more than 150 microbial whole genome sequencing projects. Our initial scope was to produce draft assemblies by generating 10x sequence coverage of standard 3kb pUC18 libraries. However, after the initial ~ 20 projects we quickly saw the holes with this methodology and reevaluated our strategy. In an effort to better span repeats and link contigs we now utilize a three library approach for all whole genome shotgun projects we undertake. We generate a 3kb high-copy pUC18 library to 4x sequence coverage, a 8kb low-copy pMCL200 library to 4x sequence coverage, and a 40kb single-copy pCC1FOS library to 15x clone coverage. A study is currently underway to evaluate the merits of increasing our 40kb fosmid library from 15x to 30x clone coverage.

Here, we will describe the construction, sequencing, and analysis of our 3 library approach with the aspiration of generating a better and more usable finished microbial whole genome assembly. The libraries were constructed from randomly sheared whole genomic DNA that was size selected and cloned into pUC18 at 3kb, pMCL200 at 8kb, and pCC1FOS at 40kb. To date, more than 130 microbial genomes have been successfully cloned using this strategy. Both the 3kb and 8kb libraries can be robustly sequenced in our high-throughput production process utilizing our streamlined Rolling Circle Amplification (TempliPhi, GE) DNA preparation method. The 40kb fosmids, on the other hand, currently pass through our production line at a slightly slower pace utilizing a 96 well bead DNA preparation method (SprintPrep, Agencourt). Analysis to date of a wide array of microbial projects indicates our cloning strategy has been successful in spanning many repeat regions and producing longer-range contiguity. These assemblies reduce the amount of finishing required to complete the genome sequence. Specific protocols and results will be described and additional details are available upon request.

Specific protocols and results are available upon request by contacting Chris Detter at (detter2@llnl.gov) or from our web site: (www.jgi.doe.gov).

***** Other related posters from the JGI on display.**

8 The US DOE Joint Genome Institute Microbial Genome Program, by Alla Lapidus and presented by Kerrie Barry.

#18 Quality Control of JGI Microbial Sequencing Projects, by Alex Copeland and Kerrie Barry.

REFERENCES

•Detter JC, Jett JM, Lucas SM, Dalin E, Arellano AR, Wang M, Nelson JR, Chapman J, Lou Y, Rokhsar D, Hawkins TL, Richardson PM. (2002). Isothermal strand displacement amplification applications for high-throughput genomics. *Genomics* 80(6):691-698.

•Dehal P, *et.al.* (2002). The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins. *Science* 298(5601):2157-2167.

•Aparicio S, *et.al.* (2002). Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. *Science* 297(5585):1301-1310.

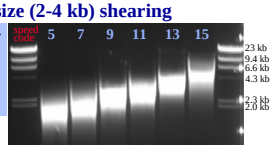
•Hawkins TL, Detter JC, Richardson PM. (2002). Whole genome amplification – applications and advances. *Current Opinion in Biotechnology* 13(1):65-67.

•Elkin C, Kapur H, Smith T, Humphries D, Pollard M, Hammon N, Hawkins T. (2002). Magnetic bead purification of labeled DNA fragments for high-throughput capillary electrophoresis sequencing. *Biotechniques*. 32(6):1296-1302.

Goal: Generate randomly sheared 3 kb shotgun libraries from whole genomic DNA via cloning into a high copy vector for high throughput end-sequencing.

Small size (2-4 kb) shearing

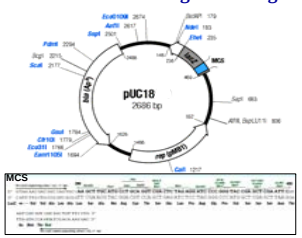
- GeneMachine Hydroshear
- Use assembly # HSA-025
- We usually use speed code 13 to obtain 3-4 kb fragments for subcloning.



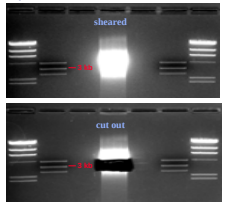
- **Name:** pUC18 (2.7kb)
- **Replicon:** pMB1
- **Purpose:** High Copy (~500/cell)
- **Selectable Marker:** Amp
- **Color Selection:** lacZ
- **Cloning Site:** SmaI

- Shear 3-5ug (HS 13)
- End-Repair
- Gel purify & size select both 2-3kb & 3-4kb
- Purify from gel (Qiagen)
- Blunt-end ligate into pUC18 at SmaI, O/N with PEG
- Phenol extract and EtOH ppt.
- Transform into electromax DH10B
- PCR QC, use modified high GC protocol if necessary

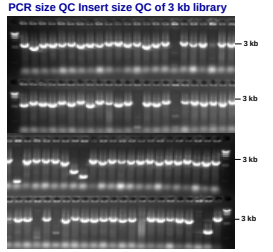
General 3 kb shotgun cloning



Gel purification and size-selection of insert



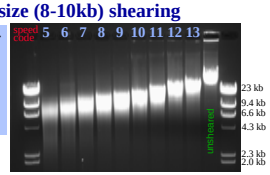
PCR size QC Insert size QC of 3 kb library



Goal: Generate randomly sheared 8 kb shotgun libraries from whole genomic DNA via cloning into a low copy vector for high throughput end-sequencing.

Medium size (8-10kb) shearing

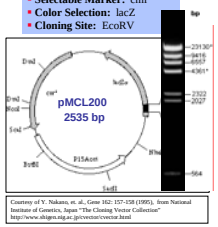
- GeneMachine Hydroshear
- Use Large assembly Catalogue # HSH 204007
- We usually use speed code 9 to obtain 8-10 kb fragments for subcloning.



Optimized 8-10kb shotgun cloning using pMCL200

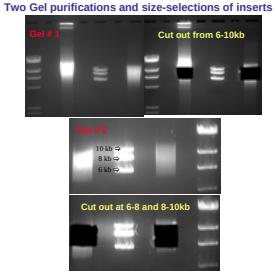
- Shear 5-10ug DNA (hydroshear, large assembly, setting #9)
- Gel purify & size select from 6-10kb
- End-Repair
- Gel purify & size select both 6-8kb and 8-10kb
- Blunt-end ligate O/N into pMCL200 w/PEG
- Phenol extract and EtOH ppt.
- Transform into eDH10B
- Size QC by long format PCR
- If high GC content, use modified high GC PCR protocol or size QC by RE digest w/ Bam HI & Hind III.

- **Name:** pMCL200 (2.5kb)
- **Replicon:** p15A
- **Purpose:** Low Copy (10/cell)
- **Selectable Marker:** cml
- **Color Selection:** lacZ
- **Cloning Site:** EcoRV

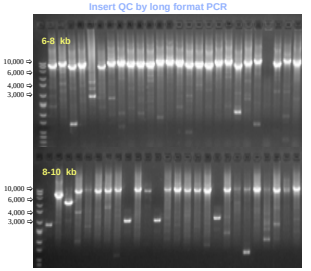


Courtesy of Y. Nakano, et. al., Gene 162: 157-158 (1995), from National Institute of Genetics, Japan "The Cloning Vector Collection" http://www.shigen.nig.ac.jp/cvector/cvector.html

Two Gel purifications and size-selections of inserts



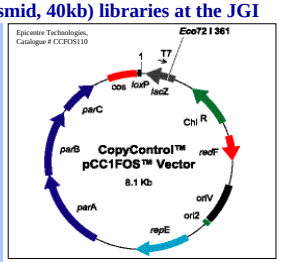
Insert size QC of 6-8 kb and 8-10kb pMCL200 library



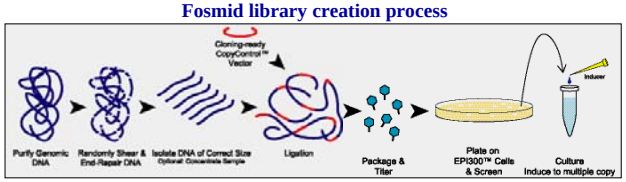
Goal: Generate randomly sheared 40 kb shotgun libraries from whole genomic DNA via cloning into a single copy vector for medium throughput end-sequencing.

Large insert (fosmid, 40kb) libraries at the JGI

- Shear 20ug (Hamilton Syringe) or (Hydroshear w/ large assembly SC#17)
- End-Repair
- Gel purify O/N & size select to ~40kb
- Phenol extract & ETOH ppt DNA
- Blunt-end ligate 40kb DNA into pCC1FOS
- Package, plate, & pick colonies
- Isolate DNA & Sequence



Fosmid library creation process



- Recent experiments have focused on developing a high-throughput fosmid DNA isolation process.
- The system we implement must be automateable and scaleable for use in a high-throughput environment.

Goal: Test the effect of assembling a whole microbial genome using end-sequences from libraries composed of three different sized inserts.

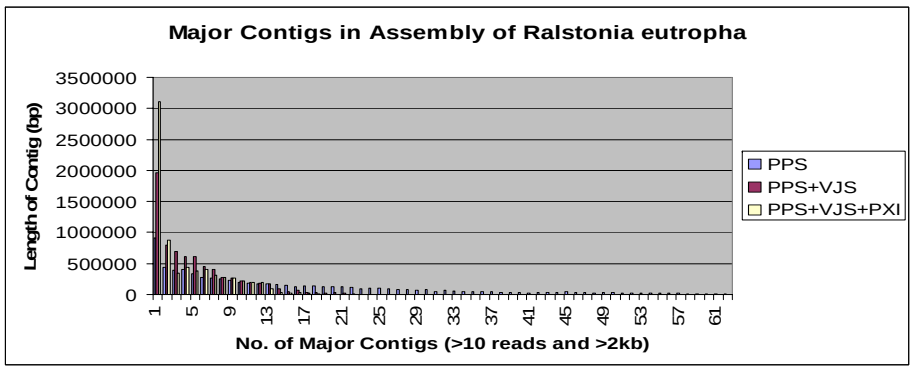
- We made 3 different sized libraries; a 3 kb in pUC18, an 8 kb in pMCL200, and a 40 kb in pCC1fos.
- Each library was sequenced to a depth shown in the table to the right.
- Phrap assemblies were performed using standard microbe assembly parameters.
- Major contigs are defined as more than 10 reads and more than 2 kb in length.

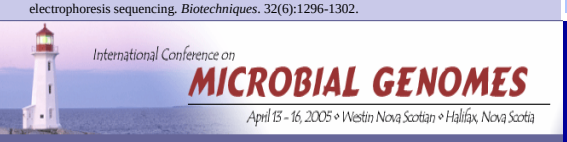
- Based on our analysis, adding reads from both 8 kb and 40 kb libraries reduces gaps and increased long range contiguity of the assembly.

LIBRARIES	Total Lanes	Pass Rate	% Vector	Q20(MB)	Coverage*
PPS (3kb)	128160	90.5%	0.8%	69.4	9.4X
VJS (8kb)	30624	96.4%	2.0%	19.4	2.6X
PXI (40kb)	22896	61.8%	0.1%	5.9	0.8X

***Ralstonia eutropha's estimated genome size is 7.4 Mb**

Assemblies	Major Contigs	Largest
PPS	62	915 Kb
PPS+VJS	26	1.96 Mb
PPS+VJS+PXI	22	3.10 Mb





US DOE Joint Genome Institute

This work was performed under the auspices of the U.S. Department of Energy, Office of Biological and Environmental Research, by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC03-76SF00098 and Los Alamos National Laboratory under contract No. W-7405-ENG-36.